Rolf H. Reichle[1,2], Randal D. Koster[2] and Sarith P. P. Mahanama[1,2]
Reichle

*Bias correction of satellite soil moisture and assimilation into the NASA Catchment land surface model

[1]Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, MD 21250, USA
[2]NASA-GSFC, Code 900.3, Greenbelt, MD 20771, USA

Surface soil moisture data from different sources (satellite retrievals, ground measurements, and land model integrations of observed meteorological forcing data) have been shown to contain consistent and useful information in their seasonal cycle and anomaly signals even though they typically exhibit very different mean values and variability. At the global scale, in particular, it is currently impossible to determine which soil moisture climatology is more correct. The biases pose a severe obstacle to exploiting the useful information contained in satellite retrievals of soil moisture in a data assimilation algorithm. A simple method of bias removal is to match the cumulative distribution functions (cdf) of the satellite and model data. Cdf estimation typically requires a long data record. By using spatial averaging with a 2 degree moving window we can obtain statistics based on a one-year satellite record that are a good approximation of the desired local statistics of a long time series. This key property opens up the possibility for operational use of current and future soil moisture satellite data.

1 Introduction and Approach

Accurate knowledge of the state of the land surface is important for many applications. For example, there is increasing evidence that accurate land initialization contributes to skill in subseasonal climate forecasts of summer mid-latitude precipitation and air temperature (Koster et al., 2003; Koster et al., 2004). Our ability to accurately characterize global soil moisture fields relies on (i) retrievals of surface soil moisture from satellite, and (ii) land surface models that integrate meteorological forcing data (such as precipitation and radiation from observations or atmospheric data assimilation) and land surface parameters (such as soil hydraulic or vegetation properties). It has long been argued that a land data assimilation system that merges these two sources of information will improve our knowledge of the state of the land surface. Such a data assimilation system must, however, address severe biases that have been identified in surface soil moisture.

(Reichle et al., 2004) show that the time series mean and variability of surface soil moisture from satellite retrievals and model integrations differ substantially, and neither agrees better with the sparse ground measurements that are available. It is in fact impossible at this time to determine a "correct" global surface soil moisture climatology towards which satellite and model data could be corrected in a data assimilation system. Rather, we are limited to removing biases between the satellite retrievals and model soil moisture by ensuring statistical consistency between the two data sources. An obvious method for doing so is to match the cumulative distribution functions (cdf's) of the satellite retrievals and model soil moisture. Similar cdf matching techniques have been used successfully for example to establish reflectivity-rainfall relationships for calibration of radar or satellite observations of precipitation (Atlas et al., 1990; Anagnostou et al., 1999). Cdf matching is conceptually straightforward, but the need to estimate the cdf of satellite retrievals of soil moisture is difficult in practice because of the limited availability of such data.

The historic Scanning Multichannel Microwave Radiometer (SMMR) offers a unique record of almost nine consecutive years of passive C-band (6.63 GHz) observations. A record of such length permits estimation of the statistics with reasonable accuracy. On the other hand, current and planned soil moisture sensors are designed for shorter observing periods of three to six years. These include the currently operational C-band Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), and two planned L-band (1.4 GHz) sensors, the Soil Moisture and Ocean Salinity (SMOS) mission, and the Hydrosphere State (HYDROS) mission. Even more importantly, it is a primary goal of these missions to provide near-real time data for operational research applications. But operational applications would be impossible if the satellite retrievals became useful only after the lifetime of the satellite. In this paper we use the nine-year SMMR record to demonstrate that temporal aggregation of SMMR data can be traded off against spatial aggregation. Robust estimation of the statistics for bias removal is accomplished using only a one-year satellite record. With our method, current and future satellite retrievals of soil moisture can be processed in near-real time using only a one-year climatology.

2 Data and Method

SMMR satellite retrievals of soil moisture are from October 1978 to August 1987 (De Jeu, 2003). The sensor's configuration on a polar-orbiting platform allowed for a maximum repeat frequency of about three to four days in mid-latitudes. Despite global coverage of the satellite, soil moisture retrievals are not available everywhere. Areas for which surface soil moisture cannot be retrieved include areas with frozen soil, mixed pixels that contain a significant fraction of surface water, and highly vegetated areas. Our land modeling system uses the state-of-the-art NASA Catchment land surface model (Ducharne et al., 2000) and surface meteorological data from (Berg et al., 2003). The surface meteorological forcing data are based on the European Centre for Medium-Range Weather Forecasting 15-year reanalysis (ERA-15) available from 1979 to 1993 and have been corrected to observed data as much as possible. Precipitation – arguably the most critical input for accurate soil moisture modeling – has been corrected primarily with a merged product of satellite and gauge data from the Global Precipitation Climatology Project (GPCP, Version 2) (Huffman et al., 1997). For further details on the SMMR retrievals and Catchment model soil moisture see (Reichle et al., 2004).

Our strategy for bias removal is to match the cdf of the satellite retrievals to the cdf of the model soil moisture by scaling the satellite retrievals. The scaled satellite retrieval $x\prime$ is given implicitly by the solution to $cdf_m(x\prime) = cdf_s(x)$, where cdf_s and cdf_m denote the cumulative distribution functions of the satellite and model soil moisture, respectively, and x is the unscaled satellite soil moisture. Note that cdf matching corrects all moments of the distribution function, subject to statistical errors that are due to a limited sample size. In practice, we can expect meaningful estimates only for the first few moments of the distribution function, and limit ourselves to analyzing the mean, standard deviation, and skewness. The key to successful bias removal via cdf matching is to identify the temporal and spatial scales at which the cdf is estimated and cdf matching is applied. Since an assimilation systems ingests instantaneous satellite retrievals at the local (or catchment) scale, statistics used for bias removal should be computed from and applied to local, instantaneous data. In fact, the complex heterogeneity of the land surface is only approximately described by the retrieval algorithm and the land model, and errors vary strongly in space. The ideal estimate of the local cdf for our purpose is thus based on the longest available data record and computed without spatial averaging.

Figure 1 shows global maps of the biases in the time series mean and standard deviation between model soil moisture and satellite retrievals. Across the globe, SMMR retrievals are typically wetter than model soil moisture, except in the eastern half of North America, northern Eurasia, and the Sahel. SMMR retrievals exhibit more variability than model soil moisture across North America, in northern Eurasia, southern Africa, and southern Australia. Elsewhere, particularly in India, SMMR retrievals are less variable in time than model soil moisture. Figure 1 demonstrates that there are severe biases between satellite retrievals and model soil moisture. Most importantly, these biases are not uniform but spatially distributed in complex patterns and are on the order of the dynamic range of the signal.

Next, we estimated the cdf based on various subsets of the full dataset of SMMR retrievals. In order to control statistical noise in the cdf estimate, we spatially aggregate the data: At any given location, observations of neighboring catchments that are within a given distance are also used to compute the statistics at the given location. In other words, we apply a moving spatial window to the computation of the statistics and implicitly assume that some degree of ergodicity is present in the data. We then use this approximate estimate of the cdf to solve $cdf_m(x\prime) = cdf_s(x)$ and obtain scaled SMMR retrievals whose statistics are again compared to those of the model soil moisture.

3 Results

When the cdf that is used for scaling is estimated from a subset of the SMMR retrievals, some locations will inevitably have insufficient data for robust estimation of the statistics. Our cutoff criterion for estimating the cdf is that at least 100 measurements must be available. Without spatial aggregation, the cdf cannot be estimated from just one year of data anywhere. Coverage increases rapidly with increasing spatial aggregation scale, and for spatial aggregation scales of 2 degrees and larger virtually no extra data are lost. The loss of coverage must be traded off against the ergodicity error, that is the error resulting from spatially averaging when estimating the cdf. We computed relative bias reduction using various 1 and 2 year subsets of the SMMR data in combination with spatial aggregation scales ranging from 0 to 5 degrees. It is important to note that the bias is always computed for the full data set (1979–1987) after scaling all SMMR retrievals with a cdf that is based on a subset of the data.

Clearly, the smallest relative biases in the mean and standard deviation are typically achieved when the full SMMR dataset (1979–1987) is used to estimate the scaling cdf. Without spatial aggregation (ideal scenario), the bias in the mean (standard deviation; or skewness) is reduced to about 2% (10%; or 45%) of the bias before cdf matching. When the full set of data is used in combination with spatial aggregation at a scale of 5 degrees, the bias in the mean (standard deviation) is about 26% (65%) of the bias before cdf matching, while the bias in the skewness after cdf matching is worse than before cdf matching. If only a single year of SMMR retrievals is used to estimate the cdf in combination with spatial aggregation, scaling with such an approximate cdf does still considerably reduce the biases. We also find that the bias after cdf matching depends only weakly on the particular year that has been used for estimating the cdf. This is not surprising, given that the biases are much larger than the interannual variability.

Knowing the loss of spatial coverage and the increase in ergodicity error as a function of the spatial aggregation scale enables us to optimize for a particular spatial aggregation scale. Since the ergodicity error increases with increasing horizontal aggregation scale, a reasonable approach is to use the minimum spatial aggregation scale for which coverage is almost complete. In our case, this strategy suggests a spatial aggregation scale of 2 degrees. In other words, the best estimate of the cdf of all SMMR retrievals that is based only on 1979 data is obtained for a spatial aggregation scale of 2 degrees. We refer to this estimate as the approximate cdf. Finally, Figure 1 also shows global maps of the remaining biases relative to the model soil moisture after bias removal using the approximate cdf for SMMR retrievals. While there is some bias left after cdf matching, bias removal with the approximate cdf based on just one year of satellite data clearly removes most of the bias in the original data.

4 Conclusions

We use the nine-year SMMR record to demonstrate that temporal aggregation of SMMR soil moisture retrievals can be traded off against spatial aggregation. Robust estimation of the statistics for bias removal via cdf matching was accomplished using only a one-year satellite record in combination with a spatial aggregation scale of 2 degrees. This approximate scenario yields almost the same coverage as the ideal scenario where the statistics are computed from a long data record without spatial aggregation. The global average bias is reduced by 98% in the ideal scenario and by 80% in the approximate scenario when compared

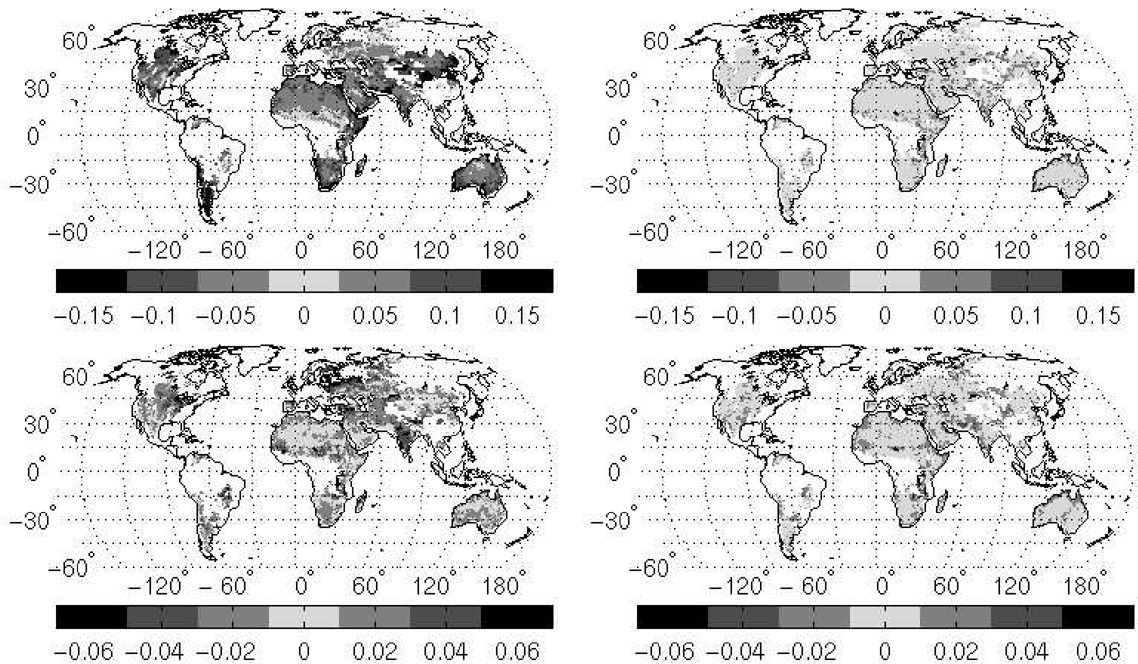

Figure 1: Difference in (Top) mean and (Bottom) standard deviation of SMMR soil moisture retrievals and model soil moisture time series (1979–1987) (Left) before and (Right) after cdf matching. Approximate cdf for scaling is estimated from 1979 SMMR data only with spatial aggregation at a scale of 2 degrees. Missing data are plotted white. Units are volumetric percent [$m^3 m^{-3}$].

to the original bias between the SMMR retrievals and model soil moisture. For the standard deviation (skewness), the ideal scenario allows bias reduction by 90% (55%) and the approximate scenario permits bias reduction by 55% (25%). With our method, current and future satellite retrievals of soil moisture can be processed in near-real time using only a one-year climatology.

An obvious assumption of the applicability of our approach to current and future sensors is that biases for AMSR-E or HYDROS retrievals relative to model soil moisture are comparable to biases encountered with SMMR retrievals. While AMSR-E and future sensors are likely to yield improved measurements of brightness temperatures when compared to SMMR, the underlying errors in the retrieval algorithm, the land surface model, and the surface meteorological forcing data are unlikely to change significantly in the near future because the retrievals used here are based on a state-of-the-art algorithm, as is the modeling system. Therefore, our approach presents a valuable tool for the imminent operational use of AMSR-E soil moisture retrievals.

References

Anagnostou, E. N., Negri, A. J. and Adler, R. F. (1999). Statistical adjustment of satellite microwave monthly rainfall estimates over amazonia, *J. Appl. Meteorol.* **38**(11): 1590–1598.

Atlas, D., Rosenfeld, D. and Wolff, D. B. (1990). Climatologically tuned reflectivity-rain rate relations and links to area-time integrals, *J. Appl. Meteorol.* **29**(11): 1120–1135.

Berg, A. A., Famiglietti, J. S., Walker, J. P. and Houser, P. R. (2003). The impact of bias correction to reanalysis products on simulations of north american soil moisture and hydrological fluxes, *J. Geophys. Res.* **108**(D16): 4490. DOI 10.1029/2002JD003334.

De Jeu, R. A. M. (2003). *Retrieval of Land Surface Parameters using Passive Microwave Remote Sensing*, PhD dissertation, Vrije Universiteit Amsterdam (Netherlands), Dept. of Geo-Environmental Sciences.

Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M. and Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model, 2: Parameter estimation and model demonstration, *J. Geophys. Res.* **105**(20): 24823.

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B. and Schneider, U. (1997). The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Amer. Meteorol. Soc.* **78**(1): 5–20.

Koster, R. D., Suarez, M. J., Higgins, R. W. and den Dool, H. M. V. (2003). Observational evidence that soil moisture variations affect precipitation, *Geophys. Res. Lett.* **30**(5): 1241. doi:10.1029/2002GL016571.

Koster, R. D., Suarez, M. J., Liu, P., Jambor, U., Berg, A. A., Kistler, M., Reichle, R. H., Rodell, M. and Famiglietti, J. (2004). Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *J. Hydrometeorol.* . in press.

Reichle, R., Koster, R. D., Dong, J. and Berg, A. A. (2004). Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, *J. Hydrometeorol.* **5**(3): 430–442.